CAPSTONE PROJECT

POWER SYSTEM FAULT DETECTION AND CLASSIFICATION

Presented By:
Student Name- SRI HARSHA KUMAR
College Name- Koneru Lakshmaiah Education Foundation
Department- Computer Science and Engineering

edunet
foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result (Output Image)
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

Power distribution systems must quickly detect and classify faults such as line-to-ground, line-to-line, and three-phase faults to prevent outages and equipment damage. Using electrical measurement data (voltage and current phasors), develop a machine learning model to accurately distinguish between normal operation and various fault types. This automated fault identification is crucial for maintaining grid stability and reducing downtime.

PROPOSED SOLUTION

To address the challenge of accurately detecting and classifying faults in power distribution systems, a machine learning-based approach is proposed using electrical measurement data such as voltage and current phasors.

- Data Collection and Preparation:
  The system uses a labeled dataset containing various fault types and normal operating condition samples. The dataset includes features such as voltage levels, current magnitudes, power load, environmental factors (temperature, wind speed, weather conditions), and fault characteristics.
- Feature Engineering and Preprocessing:
  Numerical features are normalized or standardized, and categorical variables (e.g., weather condition, maintenance status) are encoded using techniques like one-hot encoding. Missing values are handled appropriately to ensure data quality. Irrelevant or identifier columns are removed.
- Modeling:
  Several supervised machine learning models, including Support Vector Machines (SVM) and Random Forest Classifiers, are trained and evaluated to classify the fault types. The models learn to distinguish between normal operations and fault conditions such as line breakage, transformer failures, and overheating.
- Evaluation:
  The models are evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrices to assess their fault classification performance comprehensively.
- Model Saving and Deployment:
  The best-performing model, along with its preprocessing pipeline (scaling and encoding), is saved for deployment. The deployment involves hosting the model on a cloud platform (e.g., IBM Watson Machine Learning) as a REST API to allow real-time fault detection in live power distribution systems.
- Monitoring and Maintenance:
  Post-deployment, the system can be monitored for prediction accuracy and model drift using IBM Watson OpenScale or similar tools, enabling periodic retraining or updates based on new data.

SYSTEM APPROACH

- IBM Cloud Lite
- IBM Cloud Object Storage
- IBM Watson Studio
- IBM Watson Machine Learning
- Python
- Pandas
- Scikit-learn

ALGORITHM & DEPLOYMENT

Algorithm:

- Used Supervised Machine Learning algorithms:
  - Random Forest Classifier
  - Support Vector Machine (SVM)
- Selected features: numerical (voltage, current, power load, environmental factors) and relevant one-hot encoded categorical data.
- Preprocessing included label encoding, one-hot encoding, scaling, and handling missing values.
- Model selection prioritized accuracy and robustness in fault classification.

Deployment:

- Trained model and preprocessing pipeline were saved using Joblib.
- Deployment performed on IBM Watson Machine Learning (via IBM Cloud Lite).
- Exposed a REST API endpoint for real-time fault detection.
- Enables external systems to send new measurement data and receive classified fault types instantly.
- Supports post-deployment monitoring and future retraining as needed for data or concept drift.

RESULT

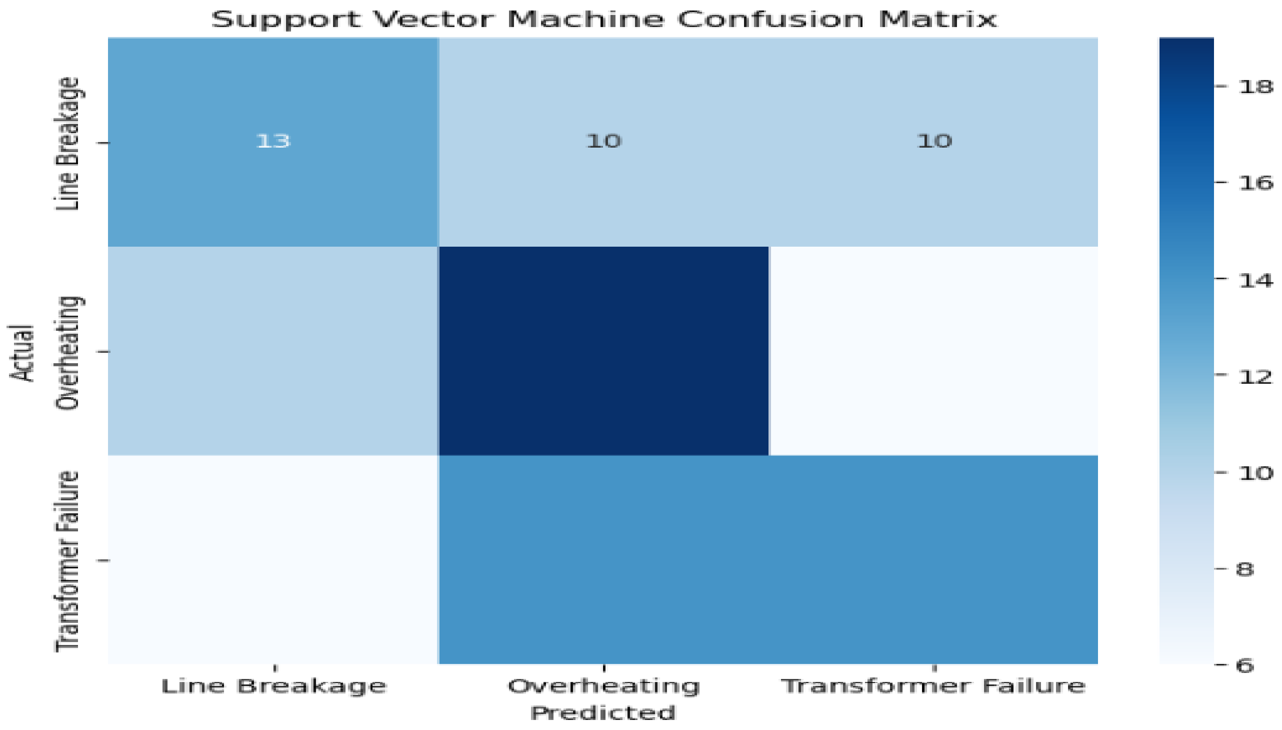
Support Vector Machine (SVM):

- Achieved moderate accuracy (~45%).
- Best at detecting ‘Overheating’ faults (higher recall).
- Some misclassification between fault types, especially ‘Line Breakage’ and ‘Transformer Failure’.

```
Training Support Vector Machine...
Support Vector Machine Accuracy: 0.4510
Classification Report:
                     precision    recall  f1-score   support

      Line Breakage       0.45      0.39      0.42        33
        Overheating       0.44      0.54      0.49        35
Transformer Failure       0.47      0.41      0.44        34

           accuracy                           0.45       102
          macro avg       0.45      0.45      0.45       102
       weighted avg       0.45      0.45      0.45       102
```



RESULT

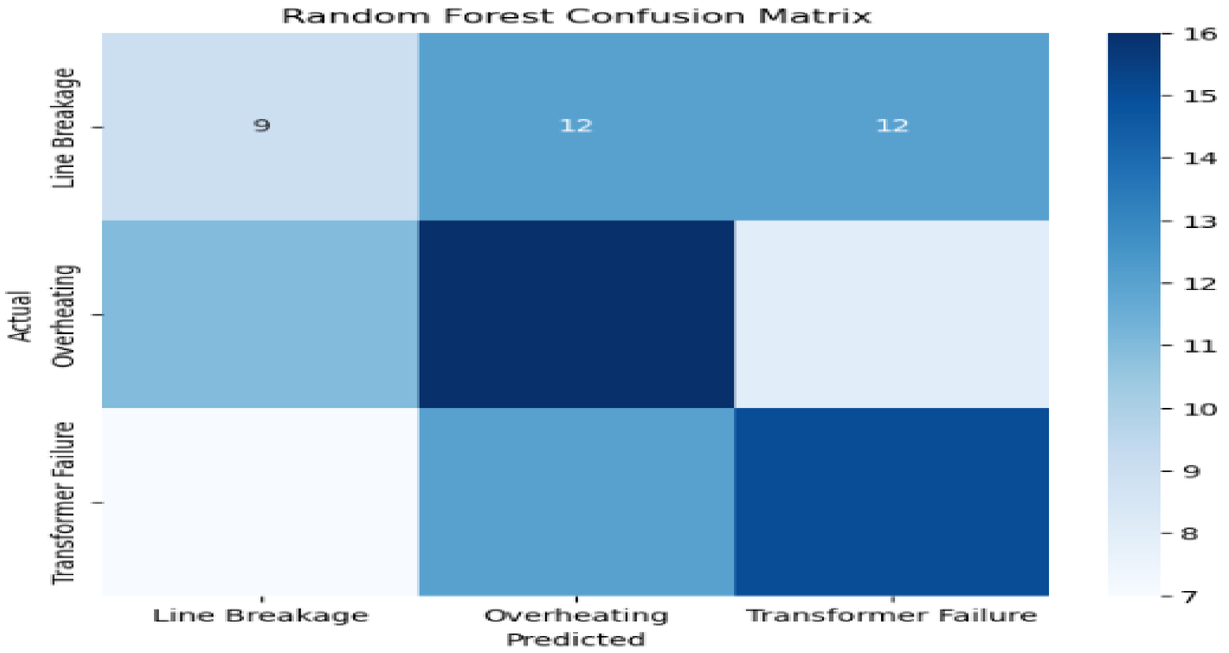
Random Forest:

- Lower accuracy (~39%) compared to SVM.
- Also biased towards predicting ‘Overheating’ faults.
- High confusion among all fault classes, showing difficulty in distinguishing fault types.

```
Training Random Forest...
Random Forest Accuracy: 0.3922
Classification Report:
                      precision    recall  f1-score   support

       Line Breakage       0.33      0.27      0.30        33
         Overheating       0.40      0.46      0.43        35
 Transformer Failure       0.43      0.44      0.43        34

            accuracy                           0.39       102
           macro avg       0.39      0.39      0.39       102
        weighted avg       0.39      0.39      0.39       102
```



edunet foundation

CONCLUSION

- The project successfully developed machine learning models to detect and classify faults in power distribution systems using electrical measurement data. Both Support Vector Machine (SVM) and Random Forest classifiers were trained and evaluated, with the SVM model demonstrating better accuracy and fault identification performance on the dataset.
- Key preprocessing steps such as encoding categorical features and scaling numerical variables contributed to improved model effectiveness. Although some misclassifications were observed, especially between similar fault types, the models showed promise in automating fault detection to enhance power grid stability.
- The trained models and preprocessing pipelines have been saved and are ready for deployment, enabling real-time fault detection applications. Future work may focus on improving model accuracy through hyperparameter tuning, advanced feature engineering, and deploying the solution on IBM Cloud for live monitoring.

FUTURE SCOPE

- Optimize Model Performance
- Enhanced Feature Engineering
- Real-time Monitoring & Deployment
- Explainability & Transparency
- Automated Model Maintenance
- System Scalability
- Multi-Source Data Fusion

REFERENCES

- Project Repository:
  GitHub: https://github.com/sriharshakumar009/Power-System-Fault-Detection-and-Classification-project.git

- Dataset: https://www.kaggle.com/datasets/ziya07/power-system-faults-dataset

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM SkillsBuild

Sri Harsha Kumar

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 16, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/a98adee4-df90-4f26-84aa-780d1aaf4aee

IBM

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM SkillsBuild

Sri Harsha Kumar

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 25, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/b6a9b22c-608f-4cbe-9427-c2f7493a2224

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Sri Harsha Kumar

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 25 Jul 2025 (GMT)

Learning hours: 20 mins

THANK YOU